

The Quasi-Holonomic Ansatz and Restricted Lattice Walks

Manuel KAUERS¹ and Doron ZEILBERGER²

Preface: A One-Line Proof of Kreweras' Quarter-Plane Walk Theorem

See: <http://www.math.rutgers.edu/~zeilberg/tokhniot/oKreweras> .

Comments: The great enumerator Germain Kreweras empirically discovered this intriguing fact, and then needed lots of pages[K], and lots of human ingenuity, to prove it. Other great enumerators, for example, Heinrich Niederhausen[N], Ira Gessel[G1], and Mireille Bousquet-Mélou[B] found other ingenious, “simpler” proofs. Yet none of them is as simple as ours! Our proof (with the generous help of our faithful computers) is “ugly” in the traditional sense, since it would be painful for a lowly human to follow all the steps. But according to *our* humble aesthetic taste, this proof is much more elegant, since it is (conceptually) *one-line*. So what if that line is rather long (a huge partial-recurrence equation satisfied by the general counting function), it takes less storage than a very low-resolution photograph.

Unrestricted Lattice Walks

Suppose that you are walking, in the d -dimensional hyper-cubic lattice Z^d , starting at the origin, and at each time-unit (you can call it a nano-second if you are a fast-walker, or a year if you are slow), you are allowed to use *any* step from a certain **finite** set of **fundamental steps**

$$S = \{(s_1, \dots, s_d)\} \quad ,$$

where each fundamental step can have *arbitrary* integer components (i.e. negative, positive, or zero).

For example, for the simple lattice (“random”) walk on the line, we have $S = \{1, -1\}$, while for the simple random walk on the two-dimensional square lattice, we have $S = \{(1, 0), (-1, 0), (0, 1), (0, -1)\}$. To cite another example, a Knight, on an infinite chessboard, is allowed any of the following eight steps:

$$S = \{(\pm 2, \pm 1), (\pm 1, \pm 2)\} \quad .$$

The quantity of interest is the $d + 1$ -variable discrete function, let's call it

$$F(m; n_1, \dots, n_d) \quad ,$$

that counts the number of ways of walking from the origin $(0, \dots, 0)$ to the point (n_1, \dots, n_d) in *exactly* m steps.

¹ Supported in part by the Austrian Science Foundation (FWF), grants SFB F1305 and P19462-N18.

² Supported in part by the USA National Science Foundation.

Often, one is interested, more specifically, in $f(m) := F(m; 0, \dots, 0)$, the number of such walks that return to the origin after m steps, and, of course we have that $g(m)$, the *total* number of walks with m steps, at the present *unrestricted* case, is trivially $|S|^m$.

It is very easy to write down the (full) *generating function* of F :

$$\tilde{F}(t; x_1, \dots, x_d) := \sum_{m=0}^{\infty} \sum_{n_1=-\infty}^{\infty} \dots \sum_{n_d=-\infty}^{\infty} F(m; n_1, \dots, n_d) t^m x_1^{n_1} \dots x_d^{n_d} \quad .$$

Indeed, the readers will have no trouble convincing themselves that

$$\tilde{F}(t; x_1, \dots, x_d) = \left[1 - t \left(\sum_{(s_1, \dots, s_d) \in S} x_1^{s_1} \dots x_d^{s_d} \right) \right]^{-1} \quad ,$$

which is a *rational* function of its variables, and it should be interpreted as a formal power series in t whose coefficients are *Laurent polynomials* in (x_1, \dots, x_d) .

It follows immediately from “general holonomic nonsense” [Z1][WZ] that $F(m; n_1, \dots, n_d)$ is *completely* holonomic, i.e. it satisfies $d + 1$ *pure* (homogeneous) linear recurrences with polynomial coefficients, one for each of its arguments. (Generically speaking. In some degenerate cases some or all of these $d + 1$ equations coincide, and one needs more equations to describe the function.).

More verbosely, there exists a positive integer L , and polynomials

$$p_0(m; n_1, \dots, n_d), p_1(m; n_1, \dots, n_d), \dots, p_L(m; n_1, \dots, n_d) \quad ,$$

such that

$$\sum_{i=0}^L p_i(m; n_1, \dots, n_d) F(m + i; n_1, \dots, n_d) = 0 \quad ,$$

for *all* $m \geq 0$ and $(n_1, \dots, n_d) \in Z^d$, and for each dimension n_i , ($i = 1 \dots d$), there exists a positive integer K_i , and polynomials $q_j^{(i)}(m; n_1, \dots, n_d)$, $j = 0 \dots K_i$, such that

$$\sum_{j=0}^{K_i} q_j^{(i)}(m; n_1, \dots, n_d) F(m; n_1, n_{i-1}, n_i + j, n_{i+1} \dots, n_d) = 0 \quad .$$

Furthermore, thanks to [MZ] (that contains, among other things, a multi-variable extension of the Almkvist-Zeilberger[AZ] algorithm), one can actually *explicitly* find these recurrences. However, for d larger than 4 and/or for large sets S , it soon becomes impractical with today’s computers.

Restricted Lattice Walk

Very often, in real life, we would like to stay in certain sub-regions of Z^d . In this case, it is no longer true that the counting function F is necessarily holonomic, as shown by Mireille Bousquet-Mélou

and Marko Petkovsek in a seminal paper [MP]. But, sometimes it is still holonomic, because of the “nice” structure of the restricted region.

For example, if the the set of steps, S , consists of the unit positive steps in d dimensions, and one is only allowed to stay in $n_1 \geq n_2 \geq \dots \geq n_d$, the famous d -dimensional *ballot* problems, that is equivalent to the number of standard Young Tableaux of shape (n_1, \dots, n_d) . Here we famously have the Young-Frobenius-MacMahon formula, that $f(n_1, \dots, n_d) := F(n_1 + \dots + n_d; n_1, \dots, n_d)$ is given by

$$f(n_1, \dots, n_d) = \prod_{1 \leq i < j \leq d} (n_i - n_j + j - i) \cdot \frac{(n_1 + \dots + n_d)!}{(n_1 + d - 1)!(n_2 + d - 2)! \dots (n_d)!} \quad ,$$

that immediately implies that not only is it holonomic, but the relevant recurrences for f are *first-order* in each of its variables, since f is expressible in *closed-form*.

There are other examples, even allowing negative steps, where one still stays in the holonomic realm, see for example [GZ]. This happens because if you put mirrors on the bounding hyper-planes, the group generated by the reflections is finite (the so-called Weyl, or Coxeter group), and the set of steps is invariant under that group.

Kreweras’ Walks

But things start to get complicated very soon. Consider the following set of **three** steps

$$S = \{(-1, 0), (0, -1), (1, 1)\} \quad ,$$

walking in **two** dimensions, and staying in the **positive quadrant**, i.e. one must stay in the region $\{(n_1, n_2) \mid n_1 \geq 0, n_2 \geq 0\}$.

Obviously, $F(m; n_1, n_2)$, defined for $m \geq 0, n_1 \geq -1, n_2 \geq -1$, satisfies the following simple recurrence

$$F(m; n_1, n_2) = F(m - 1; n_1 + 1, n_2) + F(m - 1; n_1, n_2 + 1) + F(m - 1; n_1 - 1, n_2 - 1) \quad ,$$

(whenever $m \geq 1$ and $n_1, n_2 \geq 0$),

subject to the **initial condition**:

$$F(0; n_1, n_2) = \begin{cases} 1, & \text{if } (n_1, n_2) = (0, 0) \text{ ,} \\ 0, & \text{otherwise .} \end{cases}$$

and the **boundary conditions**:

$$F(m; n_1, n_2) = 0 \text{ if } n_1 = -1 \text{ or } n_2 = -1 \quad .$$

Surprisingly, $F(m; 0, 0)$ is **closed-form**. In a classic paper, Germain Kreweras[K] proved that

$$F(3n; 0, 0) = \frac{4^n}{(n + 1)(2n + 1)} \binom{3n}{n} \quad ,$$

(of course $F(m; 0, 0) = 0$ if m is not a multiple of 3).

A naive approach would be to try and conjecture a closed-form formula, in terms of m, n_1, n_2 , for the general $F(m; n_1, n_2)$, verify that this formula obeys the above simple recurrence and the initial and boundary conditions, and finally plug-in $n_1 = 0, n_2 = 0$.

Alas, while $F(m; 0, 0)$ is almost as nice as could be, the general $F(m; n_1, n_2)$ is a huge mess, and the above approach is doomed to failure, at least if taken literally. We will later show how to rescue this simple-minded approach, by reasoning in the holonomic (or if necessary, *quasi-holonomic*) realm.

Approaches

The most successful approach so far, was to derive a **functional equation**, using combinatorial ([K]) or probabilistic ([G1]) reasoning, or the Kernel method, brought to new heights by “La Mireille” [B]. A very nice systematic study of the successes of the Kernel method, still in the quarter plane, and with **exactly three** steps, all with coordinates between -1 and 1 , was undertaken by Marni Mishna[Mi].

Ira Gessel’s Intriguing Conjecture

If the set of steps is

$$S = \{(-1, 0), (1, 0), (-1, -1), (1, 1)\} \quad ,$$

still staying in the positive quadrant ($\{(x, y) \mid x \geq 0, y \geq 0\}$), then Ira Gessel[G2] discovered empirically that (recall that $(a)_n := a(a+1) \cdots (a+n-1)$),

$$F(2n; 0, 0) = 16^n \frac{(5/6)_n (1/2)_n}{(2)_n (5/3)_n} \quad .$$

(Of course $F(2n+1; 0) = 0$). At this time of writing, as far as we know, this remains unproved. The Kernel method, so far, did not succeed, perhaps because that *now* there are **four** steps.

The Holonomic Approach

In [Z2], an empirical-yet-rigorous approach for enumerating *unrestricted* lattice paths was suggested, using the *holonomic* ansatz. This method should, at least in principle, but very possibly also in practice, succeed in doing the Kreweras problem. We now know, a posteriori, that the full generating function $\tilde{F}(t; x_1, x_2)$ for Kreweras walk is even *algebraic*, and hence *a fortiori*, holonomic. Hence, there exists, a (giant!) linear recurrence operator

$$\mathcal{P}(M, m, n_1, n_2) \quad ,$$

where M is the shift operator in m , (i.e. $Mf(m) := f(m+1)$ for any function $f(m)$), annihilating $F(m; n_1, n_2)$. It turns out that $\mathcal{P}(M, m, n_1, n_2)$ is extremely complicated, but once found, plugging-in $n_1 = 0, n_2 = 0$ gives an operator, $\mathcal{P}(M, m, 0, 0)$, annihilating $F(m; 0, 0)$.

How to prove that the empirically-derived operator does indeed annihilate F ?

Let's restrict attention to the quarter-plane. Similar reasonings apply to higher dimensions and more general regions.

Given a set of steps S , our discrete function $F(m; n_1, n_2)$ satisfies the recurrence

$$F(m; n_1, n_2) = \sum_{(s_1, s_2) \in S} F(m-1; n_1 - s_1, n_2 - s_2) \quad ,$$

which means that $F(m; n_1, n_2)$ is **annihilated** by the linear recurrence operator with **constant** coefficients

$$\mathcal{Q} = 1 - M^{-1} \left(\sum_{(s_1, s_2) \in S} N_1^{-s_1} N_2^{-s_2} \right)$$

We want to prove that $\mathcal{Q}F = 0$ plus the obvious initial and boundary conditions, imply that $\mathcal{P}F = 0$.

Let's call an operator *good* if it only contains non-negative exponents of the shift operators N_1, N_2 . For example $1 - M^{-1}N_1^2 - M^{-2}N_2$ is good but $1 - M^{-1}N_1^{-1}$ is not.

By taking commutators, or otherwise, we find, calling $\mathcal{P}_0 = \mathcal{P}$, a **sequence of good operators**, $\mathcal{R}_0(m, n_1, n_2, M, N_1, N_2), \dots, \mathcal{R}_d(m, n_1, n_2, M, N_1, N_2)$, and operators $\mathcal{P}_1(m, n_1, n_2, M, N_1, N_2), \dots, \mathcal{P}_d(m, n_1, n_2, M, N_1, N_2)$, of lower-and-lower degrees such that

$$\mathcal{Q}\mathcal{P}_0 = \mathcal{R}_0\mathcal{Q} + \mathcal{P}_1 \quad ,$$

$$\mathcal{Q}\mathcal{P}_1 = \mathcal{R}_1\mathcal{Q} + \mathcal{P}_2 \quad ,$$

...

$$\mathcal{Q}\mathcal{P}_d = \mathcal{R}_d\mathcal{Q} + \mathcal{P}_{d+1} \quad ,$$

with $\mathcal{P}_{d+1} = 0$. Since \mathcal{R}_d is "good", and since $\mathcal{Q}F = 0$, we have that $\mathcal{R}_d\mathcal{Q}F = 0$ and hence $\mathcal{Q}[\mathcal{P}_dF] = 0$. Then check that the boundary conditions for \mathcal{P}_dF is the same and the initial condition is identically 0 to deduce that $\mathcal{P}_dF = 0$. By backwards induction we (or rather our computer, it can all be mechanized) in turn, proves $\mathcal{P}_{d-1}F = 0, \mathcal{P}_{d-2}F = 0, \dots, \mathcal{P}_0F = 0$.

Note that if you don't insist that the \mathcal{R}_i 's are "good" one can always take $\mathcal{R}_i = \mathcal{P}_i$, and \mathcal{P}_{i+1} is simply the commutator of \mathcal{Q} and \mathcal{P}_i , for $i = 0, 1, \dots, d$. Since \mathcal{Q} is constant-coefficients, taking commutators with any operator with polynomial coefficients, always decreases the degree of the polynomial coefficients, so if the degree is d , eventually, after $d+1$ iterations, we get that $\mathcal{P}_{d+1} = 0$. If we want the \mathcal{R}_i to be good, we have to adjust things to be good.

In fact, for the lattice-paths-counting problems treated here, with the time variable m , starting at time $m = 0$ at the origin, it is not really necessary to demand that the \mathcal{R}_i be "good". We can

consider the function F to be defined everywhere, with 0 at the forbidden region, and rephrase that $QF = \delta(n_1, n_2)$ when $m = 0$, where $\delta(n_1, n_2)$ is the discrete delta function that is 1 at the origin and 0 elsewhere.

The Quasi-Holonomic Approach

For the sake of exposition, let's stay in the plane (analogous reasoning applies in general), and let n denote discrete time and (a, b) discrete space.

As mentioned above, Mireille Bousquet-Mélou and Marko Petkovsek proved that it is not always true, for arbitrary steps and arbitrary boundaries, that the counting function is holonomic. It is probably usually false, and the holonomicity of the Kreweras walks, and the few other cases in which it may hold, are just flukes (or follow from other considerations).

But who cares about holonomicity? Maybe it is asking way too much. Suppose, like, in the case of Gessel's conjecture mentioned above, $F(n; 0, 0)$ turns out to be holonomic in the *single* variable n .

If $F(n; a, b)$ is holonomic in *all* its arguments, then there exist *three* independent, pure recurrence operators

$$\mathcal{P}_1(n, a, b, N) \quad , \quad \mathcal{P}_2(n, a, b, A) \quad , \quad \mathcal{P}_3(n, a, b, B) \quad ,$$

annihilating F . In particular, $\mathcal{P}_1(n, 0, 0, N)$ would give us the desired operator.

But, very likely, $F(n; a, b)$ is *not* holonomic, and even if it is, like in Kreweras' case, $\mathcal{P}_1(n, a, b, N)$ is too big. What do we do now? Something much more modest would do the job!

All we need is *one* linear recurrence operator with polynomial coefficients of the form

$$\mathcal{R}(a, b, n, A, B, N) = \mathcal{R}_0(n, N) + a\mathcal{R}_1(a, b, n, A, B, N) + b\mathcal{R}_2(b, n, A, B, N) \quad ,$$

with $\mathcal{R}_0 \neq 0$.

Once found, empirically, one can prove that it annihilates our counting function $F(n, a, b)$ as above, by constructing a sequence of operators (by taking commutators, and possibly tweaking to get good operators). Once \mathcal{R} is found, and proved to indeed annihilate $F(n, a, b)$ (all of which should be done completely automatically by the computer), all we have to do is plug-in $a = 0, b = 0$ in

$$\mathcal{R}(a, b, n, A, B, N)F(n, a, b) = 0 \quad ,$$

and get that

$$\mathcal{R}_0(n, N)F(n, 0, 0) = 0 \quad .$$

QED.

Our “one-line” proof of Kreweras' theorem, mentioned in the preface, used this quasi-holonomic ansatz, even though, in this case, it is known that the counting function is holonomic. Staying within the holonomic ansatz would have made the “one-line” yet longer and its computation yet

slower. (A holonomic operator $\mathcal{R}(a, b, n, N)$ for the Kreweras walks is, for comparison, also available on the website of this article: `krewerasComplete.m`).

Analogues to Kreweras' theorem can be found effortlessly for all the eleven walks that Mishna[Mi] has isolated as being essentially different. The results are as follows:

	step set	number of closed paths
1	$\{(0, 1), (1, 1), (0, 1)\}$	$f(n, 0, 0) = 0$
2	$\{(0, 1), (1, 1), (-1, -1)\}$	$f(2n, 0, 0) = \frac{4^n (1/2)_n}{(1)_{n+1}}$
3	$\{(0, 1), (1, 1), (1, -1)\}$	$f(n, 0, 0) = 0$
4	$\{(0, 1), (0, -1), (1, -1)\}$	$f(2n, 0, 0) = \frac{4^n (1/2)_n}{(1)_{n+1}}$
5	$\{(-1, 0), (0, -1), (1, 1)\}$	$f(3n, 0, 0) = \frac{2 \cdot 27^{n-1} (4/3)_{n-1} (5/3)_{n-1}}{(5/2)_{n-1} (3)_{n-1}}$
6	$\{(0, 1), (1, 0), (-1, -1)\}$	$f(3n, 0, 0) = \frac{2 \cdot 27^{n-1} (4/3)_{n-1} (5/3)_{n-1}}{(5/2)_{n-1} (3)_{n-1}}$
7	$\{(-1, 0), (0, 1), (1, -1)\}$	$f(3n, 0, 0) = \frac{27^{n-1} (4/3)_{n-1} (5/3)_{n-1}}{(3)_{n-1} (4)_{n-1}}$
8	$\{(0, 1), (-1, -1), (1, -1)\}$	$f(4n, 0, 0) = \frac{2 \cdot 64^{n-1} (5/4)_{n-1} (3/2)_{n-1} (7/4)_{n-1}}{(2)_{n-1} (5/2)_{n-1} (3)_{n-1}}$
9	$\{(0, -1), (1, 1), (1, -1)\}$	$f(n, 0, 0) = 0$
10	$\{(-1, 1), (0, 1), (1, -1)\}$	$f(n, 0, 0) = 0$
11	$\{(-1, 1), (1, 1), (1, -1)\}$	$f(n, 0, 0) = 0$

Computer-generated proofs for the non-zero entries can be found on the website of this article.

We have also searched for an operator $\mathcal{R}(a, b, n, A, B, N)$ that would yield a proof of Gessel's conjecture, but it has turned out that no such operator can be found whose degree in A, B, N individually is at most 8 and whose total degree in a, b, n is at most 6.

A more refined counting

Another interesting problem is as follows. Given a set of steps $S = \{S_i | i = 1 \dots r\}$, count the number of walks with exactly A_i steps of kind S_i . Now the condition that it stays in the quarter-plane (or half-line, or eighth-space, or whatever), can be expressed as walks, with *positive unit steps* in \mathcal{N}^r confined to the positive sides of certain hyperplane. For example, for Kreweras's walks, if $f(a, b, c)$ is the number of walks using a steps of kind $(-1, -1)$, b steps of kind $(1, 0)$ and c steps of kind $(0, 1)$, we are counting walks from the origin to (a, b, c) staying in $c \geq a$ and $b \geq a$. Then $f(n, n, n)$ is what we called above $F(3n; 0, 0)$.

To get the quantity of interest in Gessel's conjecture, we need to compute

$$G(n) := \sum_{a=0}^n f(a, a, n-a, n-a) \quad .$$

Even though $f(a, b, c, d)$ is unlikely to be holonomic, let's hope that it is quasi-holonomic enough to guarantee that $G(n)$ is holonomic in the single variable n , a fact that we already know empirically, but it would be nice to prove it.

The Maple package `WalkCarefully` counts walks this way.

Open problem (even empirically)

Is the analog of Kreweras' walk in three dimensions holonomic?

In other words does the sequence $a(n) :=$ the number of ways of walking in the positive eighth-space ($\{(x, y, z) \mid x \geq 0, y \geq 0, z \geq 0\}$), starting at the origin, walking $4n$ steps, and returning to the origin, only employing the steps

$$\{(-1, -1, -1), (1, 0, 0), (0, 1, 0), (0, 0, 1)\} \quad ,$$

a solution of a linear recurrence equation with polynomial coefficients? It does not seem so according to our computations.

What About Gessel's Problem

The *holy grail* for lattice-walk-counters, currently, is a proof of Ira Gessel's conjecture. We strongly believe that the counting function is *quasi-holonomic*, so the present approach should, *at least in principle*, prove it. But, of course, it remains to be seen whether our proverbial margin is wide enough to contain the proof.

We also strongly believe that there is a much simpler proof (in all senses of the word) of that conjecture, that requires less than 1K of memory. That simple proof would come once the *right* and *natural* ansatz to which the (restricted) counting function belongs to, will be discovered. To give an analogy, we can routinely prove that

$$\sum_{k=0}^{10000000} \binom{10000000}{k} x^k y^{10000000-k} = (x + y)^{10000000} \quad ,$$

by staying in the *polynomial* ansatz. But it would be much more efficient to first prove that

$$\sum_{k=0}^n \binom{n}{k} x^k y^{n-k} = (x + y)^n \quad ,$$

for *all* n , by working in the *holonomic* ansatz, using WZ theory, say, and then, simply, plug-in $n = 10000000$.

The hard part, of course, for which we still need humans, is to *cherchez l'ansatz*.

Maple and Mathematica Packages

This article is accompanied by four very basic Maple packages, that compute the counting functions and empirically guess recurrences.

These are:

HalfLine , OneDimWalks, QuarterPlane, WalkCarefully .

There are also Mathematica packages

Guess , Walks

All these are available from the website of this article:

<http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/quasiholo.html> .

It is hoped that these can be extended to prove, fully automatically, Gessel's conjecture, as well as make up their own conjectures and proofs for other sets of steps.

References

[AZ] G. Almkvist and D. Zeilberger, *The method of differentiating under the integral sign*, J. Symbolic Computation **10**, 571-591 (1990).

[B] M. Bousquet-Mélou, *Walks in the quarter plane: Kreweras' algebraic model*, Annals of Applied Probability **15** (2005), 1451-1491.

[BP] M. Bousquet-Mélou and M. Petkovsek, *Walks confined in a quadrant are not always D-finite*, Theor. Computer Sci. **307** (2003), 257-276.

[G1] I. Gessel, *A probabilistic method for lattice path enumeration*, J. Stat. Plann. Inference **14** (1986), 49-58.

[G2] I. Gessel , *private communication*.

[GZ] I. Gessel and D. Zeilberger, *Random Walk in a Weyl chamber*, Proc. Amer. Math. Soc. **115**, 27-31 (1992).

[K] G. Kreweras, *Sur une class de problèmes liés au treillis des partitions d' entiers*, Cahiers du B.U.R.O, **6** (1965), 5-105.

[Mi] M. Mishna, *Classifying lattice walks restricted to the quarter plane*, preprint. Available from arxiv.org (# 0611651).

[MZ] M. Mohammed and D. Zeilberger, *Multi-Variable Zeilberger and Almkvist-Zeilberger Algorithms and the Sharpening of Wilf-Zeilberger Theory*, Adv. Appl. Math. **37** (2006), 139-152.

[N] H. Niederhausen, *The ballot problem with three candidates*, European J. Combin., **1** (1980), 175-188.

[WZ] H.S. Wilf and D. Zeilberger, *An algorithmic proof theory for hypergeometric (ordinary and "q") multiset/integral identities*, Invent. Math. **108**, 575-633 (1992).

[Z1] D. Zeilberger, *A Holonomic systems approach to special functions identities*, J. of Computational and Applied Math. **32**, 321-368 (1990).

[Z2] *The Holonomic ansatz I. Foundations and Applications to Lattice Paths Counting*, Annals of Combinatorics **11** (2007), 227-239.

Manuel Kauers, Research Institute for Symbolic Computation, J. Kepler Univ. Linz, Austria.
mkauers@risc.uni-linz.ac.at .

Doron Zeilberger, Mathematics Department, Rutgers University (New Brunswick), Piscataway, NJ, USA. zeilberg@math.rutgers.edu .